LARGE-SCALE IMAGE PROCESSING WITH THE ROTSE PIPELINE FOR FOLLOW-UP OF GRAVITATIONAL WAVE EVENTS

L. K. Nuttall[1], D. J. White[2], P. J. Sutton[1], E. J. Daw[2], V. S. Dhillon[2], W. Zheng[3], C. Akerlof[3]

[1]School of Physics and Astronomy, Cardiff University, Cardiff, CF24 3AA, United Kingdom
[2]Department of Physics and Astronomy, University of Sheffield, Sheffield, S3 7RH, United Kingdom and
[3]Randall Laboratory of Physics, University of Michigan, 450 Church Street, Ann Arbor, MI, 48109-1040, USA
Draft version February 6, 2013

ABSTRACT

Electromagnetic (EM) observations of gravitational-wave (GW) sources would bring unique insights into a source which are not available from either channel alone. However EM follow-up of GW events presents new challenges. GW events will have large sky error regions, on the order of 10-100 $\deg^2$, which can be made up of many disjoint patches. When searching such large areas there is potential contamination by EM transients unrelated to the GW event. Furthermore, the characteristics of possible EM counterparts to GW events are also uncertain. It is therefore desirable to be able to assess the statistical significance of a candidate EM counterpart, which can only be done by performing background studies of large data sets. Current image processing pipelines such as that used by ROTSE are not usually optimised for large-scale processing. We have automated the ROTSE image analysis, and supplemented it with a post-processing unit for candidate validation and classification. We also propose a simple ad hoc statistic for ranking candidates as more likely to be associated with the GW trigger. We demonstrate the performance of the automated pipeline and ranking statistic using archival ROTSE data. EM candidates from a randomly selected set of images are compared to a background estimated from the analysis of 102 additional sets of archival images. The pipeline's detection efficiency is computed empirically by re-analysis of the images after adding simulated optical transients that follow typical lightcurves for gamma-ray burst afterglows and kilonovae. We show that the automated pipeline rejects most background events and is sensitive to simulated transients to limiting magnitudes consistent with the limiting magnitude of the images.

Subject headings: gravitational waves – compact object mergers – EM transients

1. INTRODUCTION

Multi-wavelength and multi-channel observations of astrophysical systems can yield insights in to the system that are not available from a single waveband. For example, the detection of gamma-ray-burst (GRB) systems in the x-ray, optical and radio bands have led to the identification of host galaxies and their redshifts, in addition to tests of theoretical models (Bloom et al. 2006; Nakar 2007; Soderberg et al. 2006). Similar benefits may be expected from multi-channel follow up of systems that emit gravitational waves (GW). Some of the anticipated advantages include identifying host galaxies, improving parameter estimation of GW events, and determining the progenitors of phenomena such as short hard gamma-ray bursts (SHBs); see for example Bloom et al. (2009). The first attempts to detect electromagnetic (EM) counterparts to candidate GW events were made during the 2009-2010 science run of the LIGO and Virgo detectors (Abbott et al. 2009; Accadia et al. 2012). Details on how this search was performed are documented in Abadie et al. (2011). A number of optical telescopes were triggered by the GW detectors, one such system being ROTSE-III. Given the GW detector sensitivities at the time of the search, it is unlikely that any of those triggers represent true astrophysical events. However these joint observations are a useful exercise in preparing for the era of Advanced gravitational-wave detectors (Harry 2010; Virgo Collaboration 2009, c. 2015+), when EM follow-ups will be performed on GW triggers of astrophysical origin.

The ROTSE collaboration has a well established image processing pipeline. This pipeline makes use of astronomical image subtraction by cross-convolution, removing the need for high quality reference images, with similar computational efficiency to other image processing procedures (Yuan & Akerlof 2008). Transient identification is based on human scanning of potential candidates identified by the pipeline, and separate generation of lightcurves of the most interesting candidates. The pipeline has proven to be successful in finding supernovae as well as GRB afterglows etc. (Rykoff et al. 2005; Quimby et al. 2007; Rykoff et al. 2009). However, the detection of optical transients associated to GW triggers presents new challenges, in particular the need to process large numbers of images to cover a typical GW error region, and the ability to assign a quantitative false alarm probability on any detected optical transient. It is therefore essential that we have an automated image processing pipeline, where large numbers of images can be processed.

In this paper we present modifications made to the ROTSE pipeline to allow the processing of large numbers of images with automated detection and tentative classification of transients. We evaluate the performance using archival ROTSE images, and use custom-built software to add simulated transients to images. The paper is organised as follows. In Section 2 we discuss the challenges associated with detecting an EM counterpart to a GW event. In Section 3 we give a brief summary of the ROTSE-III telescope system as well as the im-

ages ROTSE took. In Section 4 we summarise how the ROTSE image processing pipeline identifies candidates. In Section 5 we describe the modifications made to automate the pipeline, including details of how the most significant candidates are identified as well as the simulation procedure. In Sections 6 and 7 we discuss the results of processing archival images to evaluate the optical transient background, and processing simulated transients to quantify the performance of the pipeline. We conclude with some brief comments in Section 8.

2. DETECTING AN EM COUNTERPART TO A GW EVENT

Many systems which produce detectable GWs should also be observable in EM wavebands (Abadie et al. 2011). The most promising GW sources which are also expected to have EM counterparts are mergers of binary neutron stars (NS-NS) or binaries consisting of a neutron star and stellar mass black hole (NS-BH). These systems are also the favoured progenitor model for SHBs (Nakar 2007). Abadie et al. (2010) summarise predictions of the rate of detection of such systems by the Advanced LIGO detectors. Metzger & Berger (2012) review various possible EM counterparts. In addition to SHBs, these include orphan optical/radio afterglows, and supernova-like optical transients ("kilonovae") generated by the decay of heavy nuclei produced in the merger ejecta (Li & Paczynski 1998; Metzger et al. 2010). Another system which may produce detectable GWs are long gamma-ray bursts (LGRBs); see Abadie et al. (2012) for a summary of possible GW emission scenarios. There is a wealth of observational data detailing the afterglow of both SHBs and LGRBs; see, for example Kann et al. (2010, 2011). The optical kilonova transient is expected to produce an optical emission peak at magnitude 18 at one day for a source at 50 Mpc and fade over the course of a few days.

GW events which produce high-energy EM counterparts such as GRBs may be promptly identified and localised by satellites such as Swift (Gehrels et al. 2004) and Fermi (Atwood et al. 2009). However, for GW events where high-energy emission is absent, or beamed away from Earth, or where the source is outside the field of view of these satellites, the detection of an EM counterpart to a GW event will be challenging. First, sky localisation using a GW data alone will produce a large error box, typically $10 - 100 \deg^2$ (Fairhurst 2009, 2011). The field of view of one of the ROTSE-III telescopes is $\sim 3 \deg^2$, making it impractical to image the entire error region. Instead, we make use of the fact that current GW detectors had a maximum distance sensitivity of between 30-70 Mpc (for NS-NS and NS-BH binary mergers) (Abadie et al. 2010) and focus observations upon galaxies in the error region within the reach of GW detectors using the galaxy catalogue described in White et al. (2011). Despite there being hundreds of galaxies in a typical GW error box, the galaxies can be ranked according to their distance and luminosity as the most likely host from which the signal originated. Considering a typical pointing with a ROTSE-III telescope, the probability of successfully imaging the correct host galaxy is estimated at between 30%-60%, not including galaxy catalogue incompleteness (Nuttall 2010). For the Advanced GW detectors, which will have an order of magnitude larger distance reach (Harry 2010; Virgo Collaboration 2009), preliminary estimates indicate that at least ~ 10 pointings will be required to have reasonable probability of imaging the host galaxy (Nuttall 2013). Nissanke et al. (2012) also present the best strategies for identifying EM counterparts to GW mergers in the Advanced detector era.

Another complication of detecting EM counterparts to GW events is that the magnitude and decay timescale of possible EM counterparts are uncertain (Abadie et al. 2011). This uncertainty necessitates observations at both early and late times, ideally from seconds to weeks after the trigger. Combined with the large error regions associated with GW triggers, this implies the need to process many images. Given the uncertain nature of the counterpart lightcurve, the image analysis should be capable of detecting any transient that is inconsistent with typical background events (which may be real astrophysical transients unrelated to the GW trigger or image artefacts).

Finally, there has not been a confirmed detection of a GW to date, making it desirable to be able to assign a high statistical confidence in any putative EM counterpart. Analysing both 'background' images (images from pointings not associated with a GW trigger) and 'injection' images (images containing simulated transients with known lightcurves) will be vital to quantify the rate at which false transients are detected as well as the performance of the pipeline. In particular, we need to test any background rejection steps on injected transients to verify they are 'safe'. All of these factors point to the need to automate the EM image analysis (see for example Bloom et al. 2011) to allow large-scale processing and quantitative characterisation of the pipeline.

3. THE ROTSE-III TELESCOPE SYSTEM

The Robotic Optical Transient Search Experiment (ROTSE) is dedicated to rapid follow up observations of GRBs and other fast optical transients on the time scale of seconds to days. ROTSE has undergone two phases of development thus far, ROTSE-I and III. ROTSE-I consisted of a 2 x 2 array of telephoto camera lenses co-mounted on a rapid-slewing platform, located in northern New Mexico. The array was fully automated and started taking data in 1998. Observations made by ROTSE-I of GRB 990123 revealed the first detection of an optical burst occurring during the gamma-ray emission, demonstrating the value of autonomous robotic telescope systems (Kehoe et al. 1999).

The ROTSE-III telescope system came online in 2003 and consists of four 0.45m robotic reflecting telescopes located in New South Wales, Australia (ROTSE- IIIa), Texas, USA (ROTSE-IIIb), Namibia (ROTSE-IIIc) and Turkey (ROTSE-IIId). The instruments are fully automated and make use of fast optics to give a 1.85 x 1.85 degree field of view. ROTSE-III is capable of attaining 17th magnitude with a 5 second exposure and 18.5 magnitude with a 60 second exposure. If multiple images are stacked on top of one another or 'coadded' ROTSE-III can reach $\sim$19th magnitude (Smith et al. 2003).

Between September 2 and October 20 2010, ROTSE-III took over 700 images in response to 5 candidate GW triggers as part of the latest science run of the LIGO and Virgo detectors (Abadie et al. 2011). All four ROTSE telescopes were used to gather the images, which span from the first night following the event to one month

later and vary in exposure length (either 20 or 60 seconds). When a LIGO/Virgo trigger was sent to the ROTSE telescopes, typically 30 images were taken on the first night and 8 images taken on subsequent follow-up nights, per telescope, for the first ten nights following the trigger, with additional observations around nights 15 and 30. While we do not use these images in this paper, we use archival images selected with this cadence so as to characterise the automated ROTSE pipeline in conditions matching that of GW follow-up observations.

4. THE ROTSE IMAGE PROCESSING PIPELINE

4.1. *Basic features*

The ROTSE image processing pipeline (Yuan & Akerlof 2008) was developed by the ROTSE collaboration to search for transient objects in images taken with the ROTSE-III telescopes. The pipeline makes use of cross-convolution to perform image subtraction. Image subtraction is an essential tool needed to remove contributions from static sources and amplify any subtle changes. For example, without image subtraction it would be almost impossible to find a source buried within a host galaxy. In this section we give a brief summary of the pipeline; more details can be found in Yuan & Akerlof (2008).

The pipeline starts by processing images through SExtractor (Bertin & Arnouts 1996), giving a list of objects with precise stellar coordinates. These coordinates are used to compute corrections for image warping, so that the stellar objects within the image overlay as closely as possible with those in the reference image. It is essential to use an image or stacked set of images (see Section 4.2) of the same region from an uninteresting time as the reference image so that a new transient may be identified. At this point in the analysis pixels within both images which exceed the saturation level are excluded. To estimate the background as precisely as possible the background difference is found between the two images, instead of the individual background for each image separately. The sky difference map is generated by performing a pixel-by-pixel subtraction between the warped and the reference image and it is this which is subtracted from the original image. The main benefit of this sky difference map is that the final subtracted image will be background-free. This procedure is repeated for all images which are to be processed before the cross-convolution algorithm is invoked.

4.2. *Coadding*

On a typical night, two sets of four images of 60 second exposure[1] with a 30 minute cadence are taken. These images are of the same part of the sky, so that images may be stacked on top of one another or 'coadded'. Coadding increases, by about one magnitude, the limiting magnitude to which we are sensitive, allowing fainter objects to be seen without saturating the brightest objects within the image. Each four-image set is coadded, as well as the eight images taken for the night, resulting in three co-additions. These three images are then subtracted from the same reference image, and the three difference

[1] A 20 second exposure is used if the target is in the vicinity of a bright galaxy or if the moon is in a bright phase.

images processed through SExtractor to reveal the residual objects.

The ROTSE pipeline can also perform a 'non-coadded' analysis, in which just the images taken from the first night are processed without coadding to see if there are any fast transients on the hour time scale. Since the non-coadded analysis does not stack images, the images have a shallower limiting magnitude than those images which have been coadded. In this paper we present examples using the coadded method only, i.e. characterising the ability to detect transients with a characteristic timescale of a few days.

4.3. *Candidate Selection*

In the coadded analysis, we have two images made from two sets of four images (called hereafter the '4-fold images') and one image made from the coadditions of all the images taken over the night (the '8-fold image') as described in Section 4.2. Any residual objects identified in these images by the pipeline are required to fulfil certain criteria to be considered candidate transients, as detailed in Yuan (2010). First, the object must have a signal-to-noise ratio (SNR) above 2.5 in the 4-fold images and above 5 in the 8-fold image. Next, the position of the object between the 4-fold and 8-fold images must match to within 1.5 pixels for candidates with SNR < 15 and to within 1 pixel for objects with SNR > 15. The full width half maximum (FWHM) of the object must be no bigger than twice the median FWHM of the stars in the convolved reference image, as well as be within the range of one pixel. The change in flux is also checked in a circular region of diameter 8 pixels around the object. Different cuts are applied depending on whether the potential candidate corresponds to a stellar object or lies in a known galaxy. For example, if an object matches a star or an unknown object a flux change of 60% is required, whereas if the object is within 20% of the semi-major axis length from the galaxy centre, but not consistent with a core, only a 3% flux change is required (Yuan 2010).

After the potential candidates have gone through these checks, further criteria are applied should more than twenty candidates remain. So many candidates remaining may indicate that the subtraction did not work correctly, or that the image quality is poor. First source crowding is checked, wherein potential candidates are rejected if they have more than 15 other potential candidates with 250 pixels. If there are still more than 20 potential candidates remaining, objects near the edge of the image are discarded, since the edges are liable to fringing and aberrations (Yuan 2010). Again, if more than 20 potential candidates remain, the area is reduced and the process repeated until the area of the image is 800 pixels in width or there are less than 20 potential candidates remaining. In these situations it is not very likely that something of astrophysical significance will be found due to the quality of the images.

Objects which have passed all the criteria outlined above form the candidate list. In fact, several candidate lists are generated: one for each night in the coadded case, and one for each consecutive pair of images in the non-coadded case. These lists need to be combined to produce a single list of unique candidates. The vast majority ($\sim 95\%$) of these potential candidates will be image subtraction artefacts, with a minority ($\sim 2\%$) due

to known variable objects such as variable stars or asteroids. We identify and remove these known transients by comparing to the SIMBAD catalogue[2] and the Minor Planet Checker[3].

4.4. Webpages

For each candidate list the pipeline also generates a webpage such as the one shown in Figure 1. At the top of the webpage three images are shown. On the left is the coadded image for one night, in the middle is the reference image, and on the right is the subtracted image. The example subtracted image shows 6 candidates. Below this are a list of links, one for each candidate. Selecting a link (in this case the first) displays a table of sub-images for that candidate. The top left panel of this table shows the first coadded image (from images 1-4 taken on that night), the top middle shows the second coadded image (from images 5-8), and the top right shows the reference image, all zoomed in to the vicinity of the candidate. The bottom left plot shows the first subtracted image (the first coadded image minus the reference), the bottom middle shows the second subtracted image. The bottom right panel displays information about the candidate, including the right ascension, declination, magnitude, signal-to-noise, FWHM (these last three quantities are calculated by comparing the reference image with the coadded image of the entire night), motion (this is the variation in distance between the first and second coadded images in units of pixels), percentage flux change (between the coadded image of the night and the reference image) and whether a candidate has been found at these coordinates before. As well there are links to the SIMBAD catalogue, Minor Planet Checker, SDSS[4], 2MASS[5] and DSS[6] to help decide the importance of the candidate. From this information, the user manually selects candidates of interest and lightcurves for these candidates are generated. It is possible to produce two lightcurves; one which includes both the transient and background and one which subtracts the background (estimated using an annulus of inner radius ∼6 pixels and outer radius of ∼14 pixels) away producing the lightcurve for just the transient.

5. AUTOMATING THE PIPELINE

The ROTSE image processing pipeline has been used to make some significant discoveries of optical transients (Rykoff et al. 2005; Kehoe et al. 1999; Baykal et al. 2005; Gezari et al. 2009; Chatzopoulos et al. 2011). However the follow up of GW events requires processing larger numbers of images that is not feasible with a widget-based, user driven setup designed to handle one set of images at a time. For example, a series of commands in the IDL environment[7] are used to produce the various lists of candidates and their corresponding webpages. Human scanning is then required to distinguish candidates of astrophysical interest from those due to poor image

[2] http://simbad.u-strasbg.fr/simbad/
[3] http://scully.cfa.harvard.edu/cgi-bin/checkmp.cgi
[4] http://www.sdss.org/
[5] http://www.ipac.caltech.edu/2mass/
[6] http://archive.stsci.edu/cgi-bin/dssform
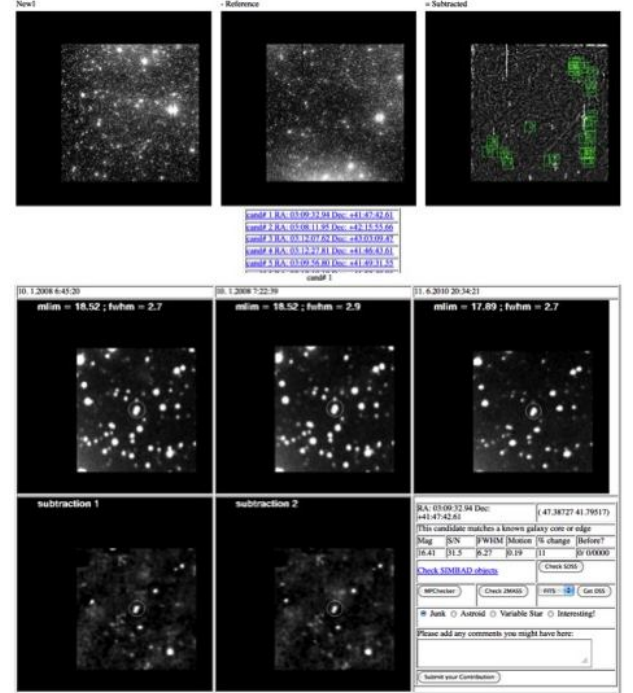[7] http://www.exelisvis.com/language/en−us/productsservices/idl.aspx

FIG. 1.— A sample ROTSE pipeline webpage, showing links to all the candidates found as well as a table displaying subimages and information for the first candidate. The full webpage displays one table for each candidate.

subtraction, those due to minor planets, etc. Further widget-based commands are then needed to produce the lightcurve of each interesting candidate. This procedure is user intensive and time consuming. However, many of these steps are algorithmic, such as checking for candidates at the same right ascension and declination across nights, and suitable for automation. We have therefore written a wrapper to the pipeline that automates the processing of large sets of images. A single command now runs the complete end-to-end pipeline: looping over image sets, finding transients, identifying transients detected across multiple nights, and generating lightcurves for all transients.

Other barriers to processing large numbers of images are the need to have an IDL license for each instance of a running pipeline, and a pipeline architecture that is designed to process only a single set of events at one time. We have also altered the pipeline architecture to automatically create separate directory structures for each set of images, allow multiple instances of the pipeline to run simultaneously without conflict. Furthermore, we have removed the need for separate IDL licenses for each instance of the pipeline by compiling the pipeline in an IDL virtual machine[8]. Only one license is required, and only at the compilation stage. Combined, the change in architecture and freedom from license restrictions enables the processing of multiple sets of images simultaneously on computer clusters. We have written scripts for large scale processing using the CONDOR/DAGMAN job management system[9] for this purpose. The automated processing is able to perform a complete analysis, iden-

[8] http://www.exelisvis.com/language/en−us/productsservices/idl/idlmodules/idlvirtualmachine.aspx
[9] http://research.cs.wisc.edu/condor

tifying candidates and generating lightcurves, within a few hours (Nuttall et al. 2012). We have verified that the automated version of the pipeline produces lists of candidates which are identical to those produced by the original manual analysis.

5.1. *Candidate Validation and Classification*

Once the automated code has produced the lightcurve information for all the potential candidates identified by the pipeline, a series of pass/fail tests are applied to each candidate. Specifically, we test whether the candidate appears on more than one night, whether its coordinates overlap with a known variable source (by querying the SIMBAD catalogue) or with an asteroid (by querying the Minor Planet Checker), and if the lightcurve of the potential candidate varies sufficiently. This last test has two components: a chi-square test to check that the candidate's lightcurve is not too flat, and a check that the lightcurve decays sufficiently 48 hours after the event took place. The multiple-night and flatness tests are very effective at rejecting non-astrophysical background, particularly image-subtraction artefacts. The decay test is seen to reduce significantly the background of astrophysical transients unrelated to the GW trigger while not rejecting simulated astrophysical transients correlated with the GW trigger (see Section 7). The specific requirement of decay after 48 hours is motivated models of EM counterparts for systems with strong GW emission, specifically kilonovae and SHB/LGRB afterglows. While there are astrophysical optical transients that do not decay on this timescale, such as supernovae (Leibundgut 2005), the expected GW emission by these sources make them less likely to produce sources of GW triggers than compact-object mergers.

We refer to these pass/fail tests collectively as the 'hard' cuts in the analysis. Any candidate which fails one or more of the hard cuts is discarded. Those which survive the hard cuts are looked at further in two ways. Firstly we see whether the candidate's coordinates overlap (to within three times the size of the major diameter) with a known galaxy. We use the Gravitational Wave Galaxy Catalogue (White et al. 2011), considering only galaxies within 50 Mpc, as this is approximately the maximum range of current GW detectors to NS-NS and NS-BH binaries (Abadie et al. 2011). Secondly, we perform a chi-square test comparing the candidate's lightcurve with several theoretical models: kilonovae, SHB afterglows, and LGRB afterglows. Candidates that fulfil any of these conditions are highlighted in the final candidate list.

The final candidate list following application of these tests typically contains fewer than 5 candidates. In order to better assess the statistical significance of any surviving candidates, we assign to each an *ad hoc* ranking statistic R defined as

$$R \equiv \sum_i \Theta(18 - m_i)(18 - m_i) \times w_i. \qquad (1)$$

Here $\Theta(x)$ is the step function, m_i is the background-subtracted magnitude of the transient in image i, and w_i

is a weight factor defined by

$$w_i = \begin{cases} 1 & t_i - t_{\mathrm{GW}} < 1\,\mathrm{day} \\ \left(1 + \log_{10} \frac{t_i - t_{\mathrm{GW}}}{1\,\mathrm{day}}\right)^{-a} & t_i - t_{\mathrm{GW}} \geq 1\,\mathrm{day} \end{cases} \qquad (2)$$

Here t_{GW} is the time of the GW trigger and t_{i} is the time of image i. The power law index a is chosen to be 3 as this is the approximate gradient of the three target theoretical lightcurves, and magnitude 18 is the approximate limit of the majority of the ROTSE images we are analysing. Candidates with magnitude $m_i > 18$ are likely to be processing artefacts, so the Θ factor ensures a rank of zero for those cases. While equation (1) is *ad hoc*, it has the desirable property of favouring brighter candidates which appear in multiple images close in time to the GW trigger.

5.2. *Simulated Transients & Detection Efficiency*

Adding simulated transients ("injecting") into the ROTSE images is key to quantifying both the detection efficiency and the magnitude limit of the pipeline. Since the processing uses image subtraction to remove the background, it is not so trivial as copying a model star and placing it somewhere else in the image. The variation in background around the transient in question has to be taken into account to realistically inject a simulated transient into the ROTSE images.

To begin, the user selects a number of real stars from the image as model stars. These stars must be sufficiently bright and isolated, so that the injection code does not take into account the flux of any unwanted stars and is able to accurately determine the point spread function (PSF) of the model star. We note that simple models for the PSF (e.g. a Gaussian) are not applicable for wide field of view images such as those from ROTSE, as the PSF varies across the image. An injection is performed by selecting a random position within 100 pixels of the model star, and selecting the distance to the source. The flux of the model star (minus the background) is scaled to follow the desired lightcurve, such as the kilonova or afterglow models discussed in Sections 2 and 5.1. The magnitude required in each image is calculated by taking into account the time between the GW trigger and the image being taken; for our tests we assume an interval of 0.5 days elapsed between the trigger time and the first image (White et al. 2012).

It is vital to inject a transient not only with the correct parameters, but also with the correct background. Simply copying a model star to a new location in the image would produce a background around the injection that is significantly higher than elsewhere in the image, as the post-injection background would comprise both the pre-injection background at that location and the background around the original model star. This could lead to the image processing pipeline identifying fainter injected transients than is realistic. We therefore scale the background around the injection by a constant amount so that the background before and after the injection is comparable; see Figure 2 for an example.

6. BACKGROUND STUDY

Assigning a statistical significance to an event identified by the pipeline as associated with a GW trigger
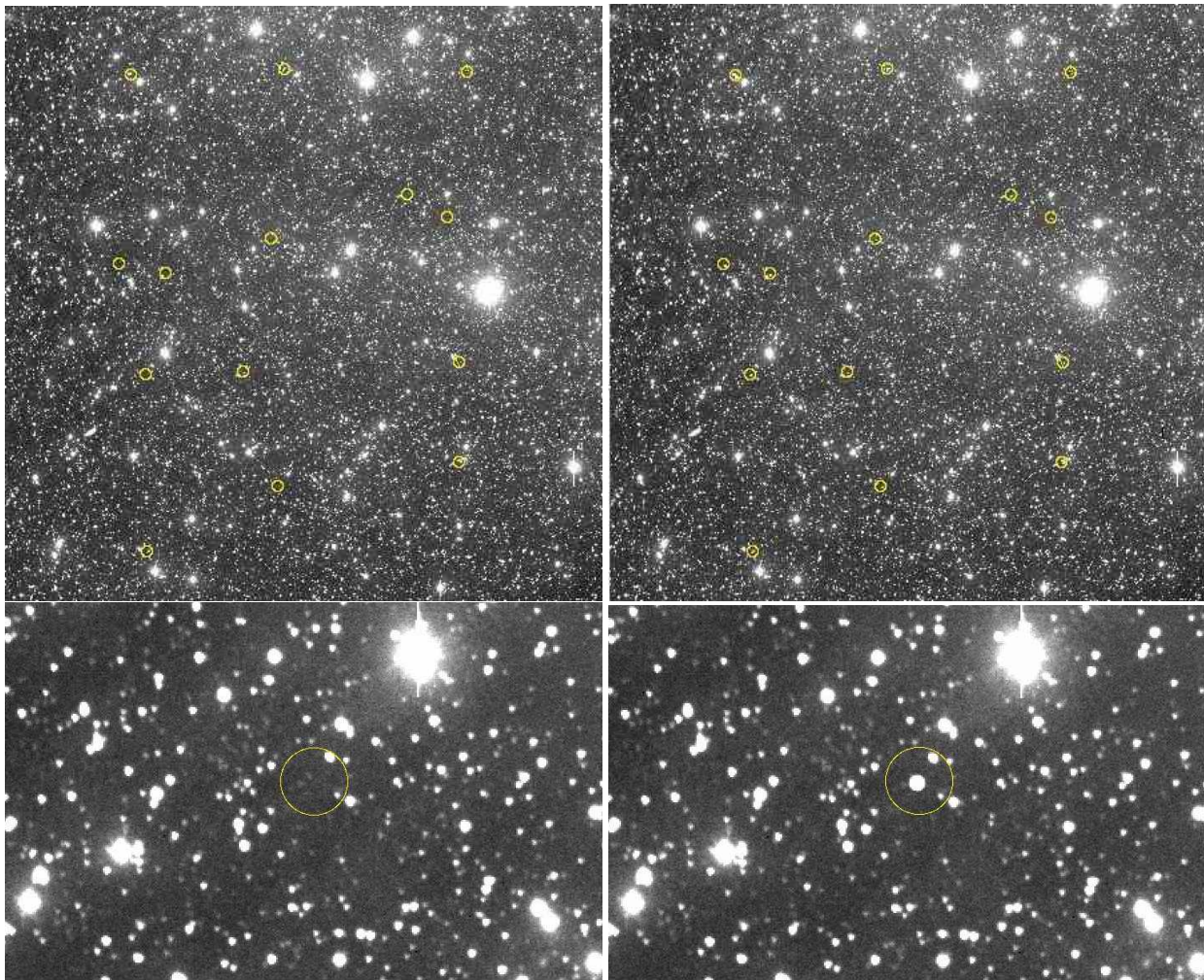
FIG. 2.— An example of injecting a number of transients into an image: (top left) original image (top right) same image with 14 injections. The regions where the injections occurred are highlighted by yellow circles in both images for comparison. (bottom) same images as top, focussed on the region around a single injection.

requires quantifying the false alarm probability. This is the probability of obtaining a similar event due to background, where for our purposes "background" includes both image-processing artefacts and real astrophysical transients that are not associated with a GW trigger. To quantify this probability we have performed a background study using archival ROTSE data. We selected at random 102 sets of images taken in response to non-GW pointings over 2 years. To better mimic a GW trigger follow-up, each set was required to have observations spanning at least a month. This yielded a total of 103 sets of images. One of these was selected at random to be our test "GW trigger", and the other 102 were used for background estimation.

The background is characterised as follows: each set of background images is processed by the automated pipeline and the highest rank R in equation (1) is found. (If a background set has no surviving candidates after the hard cuts, a rank of zero is recorded.) The distribution of highest-ranked events for our 102 background pointing sets is shown in Figure 3. We find a bi-modal distribution where approximately 80% of the pointings having a ranking statistic of less than 1 and approximately 10% have a rank greater than 11. The highest-ranked back-
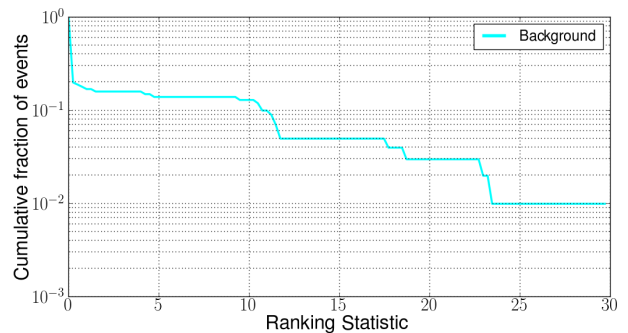

FIG. 3.— The distribution of ranking statistic R (equation (1)) for the highest-ranked transient in each of the 102 background image sets from the ROTSE archive. Image sets with no candidates surviving after the hard cuts are assigned a rank of zero. The highest-ranked background transient over the 102 sets has a rank of $R = 29.5$.

ground event has $R \sim 30$. A candidate in the GW trigger image set would therefore require $R \gtrsim 11$ ($R \gtrsim 30$) to have a false alarm probability of 0.1 (0.01) or smaller.

7. INJECTION STUDY

In order to test the robustness of the pipeline we have performed an injection study whereby transients of a

given model (either kilonova, SHB afterglow, or LGRB afterglow) are injected into the archival ROTSE images selected as our test "GW trigger". We choose 14 reference stars in the first image as our models for the injections. These reference stars are chosen as uniformly as possible so injection performance may be tested across the image. This is then repeated 10 times so that 140 injections of each model are performed at a given distance. For convenience, we choose to inject all the models at similar magnitudes, corresponding to different source distances. For example, a kilonova at 1 Mpc corresponds to a SHB afterglow at a distance between 0.2 and 8 Mpc and a LGRB afterglow between 5 and 200 Mpc. The spread of distance for the gamma-ray models reflects the uncertainty in the lightcurve observations as outlined in (Kann et al. 2010, 2011). Injection magnitudes between 8 and 17 were tested, corresponding to distances between 0.4 and 30 Mpc for the kilonova model and larger distances for the afterglow models.

Figure 4 shows the efficiency of the pipeline in finding the injections, irrespective of rank R, in terms of distance and magnitude. The distances quoted for the GRB afterglow models are best-case, using the most optimistic luminosities in Kann et al. (2010, 2011); the worst-case luminosities give distances which could be a factor 40 lower. At very close distances or low magnitudes all efficiencies suffer from saturation: the injections are so bright that their image pixels are saturated. As described in Section 4.1, the pipeline removes saturated pixels at a very early stage as they are assumed bad and not astrophysically interesting. Attempts have been made to overcome this issue by fitting each of the injection models to the data. The best-fit model is selected and used to predict the magnitude at the time of each image. For any images for which the candidate is not reported by the pipeline and for which the predicted magnitude is low enough to cause saturation, a new rank is calculated using the predicted magnitude for that time. We find that this procedure successfully retrieves transients ∼1 magnitude too bright for the unaltered pipeline, but it is not effective for even brighter (closer) transients.

Figure 4 shows that the automated pipeline achieves detection efficiencies above 50% for all models over magnitudes of approximately 9 – 14. The maximum detection efficiency is greater than 60% for all models tested. Of the 35% – 40% of injections which are not found, more than half are lost because insufficient lightcurve data could be generated. As described in Section 4.4, the pipeline generates two lightcurves for each candidate, one with background subtracted, and one without the background subtracted. Our ranking statistic R (equation 1) is based on the background-subtracted lightcurve which is highly dependent on image quality and the position of the transient in the image. If we were to only require a transient to pass the hard cuts described in Section 5.1, the peak efficiency for each model would be closer to 90%. The limiting magnitude varies across the image, with the most sensitive region at the centre and the least around the edge of the image. Therefore the efficiency of injection finding begins to decrease above magnitude 14 as they are not found in the edge regions first, until at the highest magnitudes only injections which are located in the centre of the image are identified.

The efficiency of detecting injections with a false alarm
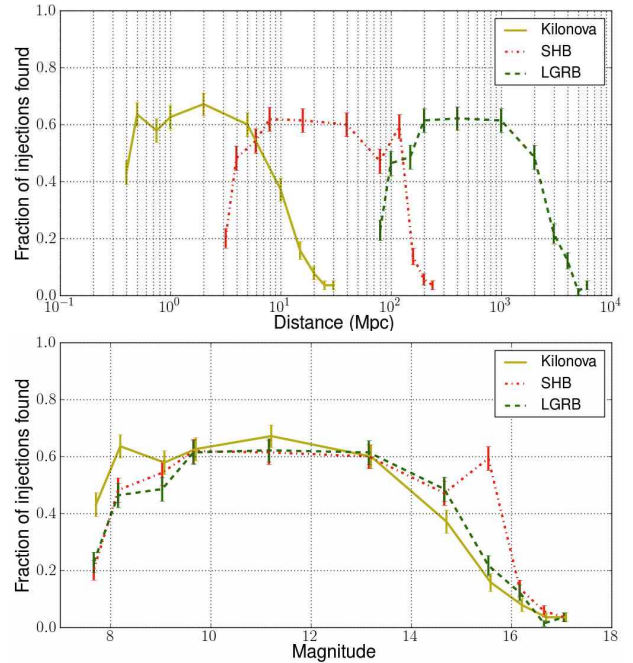

FIG. 4.— (top) Fraction of injections found by the automated pipeline, irrespective of rank, versus distance. The distances quoted for the GRB models are a best-case scenario, and could be worse by a factor of up to 40 (Kann et al. 2010, 2011). (bottom) Efficiency versus injection magnitude. All the models suffer from poor efficiency at very close distances / low magnitudes due to saturation.

probability of less than 10% (i.e., with $R \gtrsim 11$) is shown in Figure 5. The efficiencies are not as high as those found in Figure 4, with maximum efficiencies between ∼45% and ∼60% depending on the model. This would therefore suggest that all candidates which pass the hard cuts should be looked at further to see whether they are astrophysically interesting. Figure 6 shows the distribution of injections, in terms of rank, at various distances. At close distances the rank of injections is higher than the loudest candidate found in the background. At a kilonova distance of 1 Mpc the loudest injections are comparable to the loudest background event. As the distance/magnitude is increased the ranks slowly fall to much lower numbers, making them unexceptional when compared to the loudest events in the background. This again lends weight that any candidate to make the final candidate list be further investigated for significance.

8. CONCLUDING REMARKS

We have presented an autonomous pipeline for large scale processing of images taken with the ROTSE-III telescope system in addition to a ranking scheme to classify potential candidates. The ranking scheme favours a transient which is seen on multiple nights, has a bright magnitude and a decaying lightcurve. A background study of more than 100 random pointings taken from the ROTSE archives has been performed as well as an injection study of more than 4500 simulated transients added to additional archival images. Results show that the ranking scheme is good at identifying transients which are injected with magnitudes brighter than 13.5 on the second night. Injections which are brighter than magnitude 9 have a ∼50% chance of suffering from saturation. Those injected transients which are found however, tend
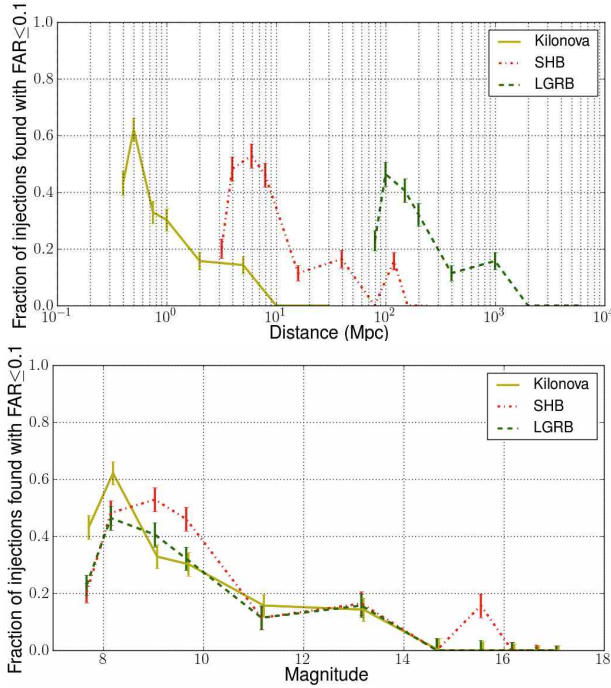
FIG. 5.— Fraction of injections found with a rank comparable to less than 10% of the background in terms of distance (top) and magnitude (bottom). The distances quoted for the gamma-ray bursts are a best case scenario, in reality they could be a factor of 40 worse.

series of pass/fail 'hard' cuts, but also on background-subtracted lightcurve data being obtained. The ability of the pipeline to produce this data for a transient depends on both the position in the image and on good image quality. The maximum injection efficiency, for all three models, is ∼60% for magnitudes 10 − 13 (in the second night's image). Requiring only that the pipeline identify an injection, regardless of lightcurve data, the maximum detection efficiency is closer to 90%. Above magnitude 13 the efficiency begins to fall as injections are not identified around the edge of the image first. This is because the limiting magnitude varies greatly over the image, with the worse limiting magnitudes at the edges and the most sensitive regions in the centre of the image. Efficiencies fall to a few percent at the largest magnitudes. All these results motivate the need for further human scanning of the handful of interesting candidates identified by the pipeline, as well as further study of ranking statistics and possible background-rejection tests.

The ability to process large sets of images in a matter of hours will be essential in the Advanced gravitational-wave detector era, where GW detections will be a regular occurrence. Although we know of some systems likely to produce both GW and EM transients, there are likely to be other sources we have not considered. During the next few years it is vital we build tools to cope with the demand to process EM data triggered from GW events.

9. ACKNOWLEDGMENTS

to have FAR much larger than the background. Those simulated transients which are recovered the best by the pipeline, with magnitudes between 9 and 13.5, tend to have FARs between 0.01 and 0.05. At greater magnitudes the injections fall within the background, and have a false alarm probability comparable to ∼20% of the background. The efficiency of identifying an injection is not only dependent on the transient passing a

REFERENCES

Abadie, J., et al. 2010, Class. Quant. Grav., 27, 173001
—. 2011, A&A, 539, A124
Abadie, J., et al. 2012, ApJ, 760, 12
Abbott, B., et al. 2009, Rept. Prog. Phys., 72, 076901
Accadia, T., et al. 2012, JINST, 7, P03012
Atwood, W. B., et al. 2009, ApJ, 697, 1071
Baykal, A., Kiziloglu, U., Kiziloglu, N., Balman, S., & Inam, S. C. 2005, Astron.Astrophys.
Bertin, E., & Arnouts, S. 1996, Astron. Astrophys. Suppl. Ser., 117, 393
Bloom, J., Richards, J., Nugent, P., Quimby, R., Kasliwal, M., et al. 2011, arXiv:1106.5491
Bloom, J. S., Prochaska, J., Pooley, D., Blake, C., Foley, R., et al. 2006, Astrophys.J., 638, 354
Bloom, J. S., et al. 2009, arXiv:0902.1527
Chatzopoulos, E., Wheeler, J. C., Vinko, J., Quimby, R., Robinson, E., et al. 2011, Astrophys.J., 729, 143
Fairhurst, S. 2009, New J. Phys., 11, 123006
—. 2011, Class.Quant.Grav., 28, 105021
Gehrels, N., et al. 2004, ApJ, 611, 1005
Gezari, S., Halpern, J., Grupe, D., Yuan, F., Quimby, R., et al. 2009, Astrophys.J., 690, 1313
Harry, G. M. 2010, Class. Quant. Grav., 27, 084006
Kann, D., Klose, S., Zhang, B., Malesani, D., Nakar, E., et al. 2010, Astrophys.J., 720, 1513
Kann, D., Klose, S., Zhang, B., Wilson, A., Butler, N., et al. 2011, Astrophys.J., 734, 96

Kehoe, R., Akerlof, C., Balsano, R., Barthelmy, S., Bloch, J., et al. 1999
Leibundgut, B. 2005, in Springer Proceedings in Physics, Vol. 99, Cosmic Explosions, ed. J.-M. Marcaide & K. W. Weiler (Springer Berlin Heidelberg), 173–182
Li, L.-X., & Paczynski, B. 1998, Astrophys.J., 507, L59
Metzger, B. D., & Berger, E. 2012, ApJ, 746, 48
Metzger, B. D., et al. 2010, arXiv:1001.5029
Nakar, E. 2007, Physics Reports, 442, 166
Nissanke, S., Kasliwal, M., & Georgieva, A. 2012
Nuttall, L., et al. 2012, J.Phys.:Conf.Ser. 363, 012033
Nuttall, L.K. & Sutton, P. 2010, Phys. Rev., D82, 102002
—. 2013, in preparation
Quimby, R. M., Wheeler, J. C., Hoflich, P., Akerlof, C. W., Brown, P. J., et al. 2007, Astrophys.J., 666, 1093
Rykoff, E., Aharonian, F., Akerlof, C., Ashley, M., Barthelmy, S., et al. 2009, Astrophys.J., 702, 489
Rykoff, E. S., Yost, S., Krimm, H., Aharonian, F., Akerlof, C., et al. 2005, Astrophys.J., 631, L121
Smith, D., Akerlof, C., Ashley, M., Casperson, D., Gisler, G., et al. 2003, AIP Conf.Proc., 662, 514
Soderberg, A. M., Berger, E., Kasliwal, M., Frail, D., Price, P., et al. 2006, Astrophys.J., 650, 261
Virgo Collaboration. 2009, Advanced Virgo Baseline Design, https://tds.ego-gw.it/itf/tds/file.php?callFile=VIR-0027A-09.pdf
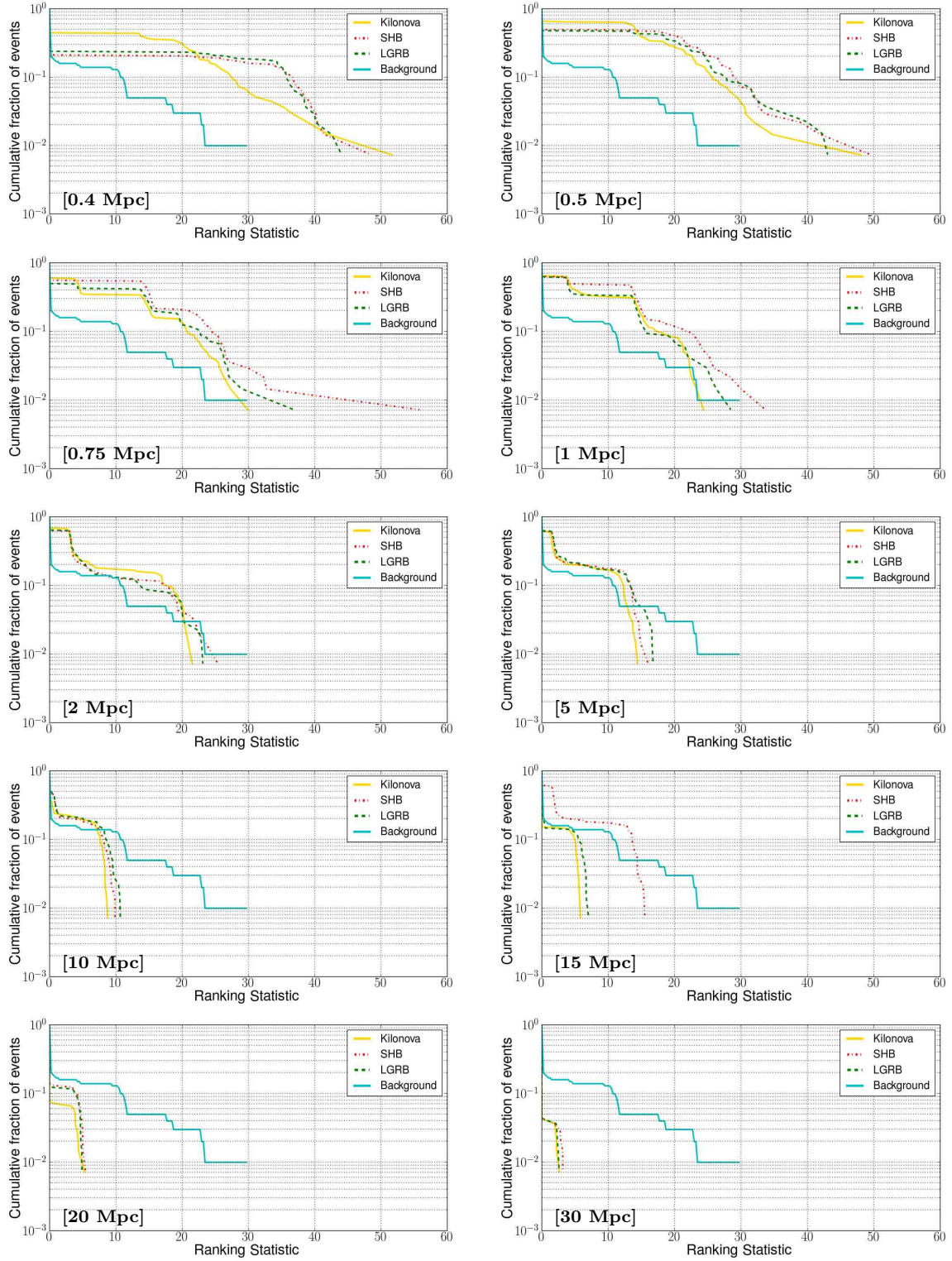
FIG. 6.— Distribution of injections compared to the background. The distance quoted is the distance at which the injections were made. In reality at a kilonova distance of 1 Mpc the SHB could be at a distance between 0.2 and 7.9 Mpc and the LGRB between 5 and 200 Mpc. This uncertainty is due to the range of luminosities in observed GRB lightcurves.

White, D., Daw, E., & Dhillon, V. 2011, Class.Quant.Grav., 28, 085016

White, D., et al. 2012, J.Phys.:Conf.Ser, 363, 012036

Yuan, F. 2010, http://hdl.handle.net/2027.42/75873

Yuan, F., & Akerlof, C. 2008, The Astrophysical Journal, 677, 808